Cloning, characterization, and expression in *Escherichia coli* of the gene coding for the CpG DNA methylase from *Spiroplasma sp.* strain MQ1(M·SssI)

Paul Renbaum, Dan Abrahamove, Abraham Fainsod, Geoffrey G.Wilson[2], Shlomo Rottem[1] and Aharon Razin

Department of Cellular Biochemistry and [1]Department of Membrane and Ultrastructure Research, The Hebrew University-Hadassah Medical School, Jerusalem, 91010 Israel and [2]New England Biolabs Inc., Beverly, MA 01915-9990, USA

Received December 12, 1989; Revised and Accepted January 16, 1990 EMBL accession no. X17195

ABSTRACT

We describe here the cloning, characterization and expression in *E. coli* of the gene coding for a DNA methylase from *Spiroplasma* sp. strain MQ1 (M·SssI). This enzyme methylates completely and exclusively CpG sequences (1). The Spiroplasma gene was transcribed in *E. coli* using its own promoter. Translation of the entire message required the use of an opal suppressor, suggesting that UGA triplets code for tryptophan in *Spiroplasma*. Sequence analysis of the gene revealed several UGA triplets, in a 1158 bp long open reading frame. The deduced amino acid sequence revealed in M·SssI all common domains characteristic of bacterial cytosine DNA methylases. The putative sequence recognition domain of M·SssI showed no obvious similarities with that of the mouse DNA methylase, in spite of their common sequence specificity. The cloned enzyme methylated exclusively CpG sequences both *in vivo* and *in vitro*. In contrast to the mammalian enzyme which is primarily a maintenance methylase, M·SssI displayed *de novo* methylase activity, characteristic of prokaryotic cytosine DNA methylases.

INTRODUCTION

Mammalian DNA is methylated almost exclusively at CpG sequences. Over the last decade we have witnessed growing interest in this phenomenon. Our interest in CpG methylation in higher eukaryotes stems from the continuously accumulating data suggesting that this methylation is associated with the regulation of gene activity (2). One line of evidence suggesting the involvement of DNA methylation in gene silencing comes from experiments using *in vitro* methylated gene sequences to study their capacity to function *in vivo* (3).

All eukaryotic DNA methylases isolated to date have proven to be maintenance enzymes methylating hemimethylated DNA 10-100 fold more efficiently than unmethylated duplex DNA (4). For that reason, use of eukaryotic enzymes to methylate unmethylated DNA *in vitro* has not been practical. As an alternative, HpaII and HhaI methylases methylating CCGG and GCGC sequences, respectively, have been used to methylate DNA *in vitro*. However, this methylation is limited to the recognition sites of each enzyme representing only 6% of the CpG sequences.

While studying DNA methylation among members of the family *Spiroplasmatacea* (belonging to the class *Mollicutes*, the wall-less bacteria) we discovered that all CpG residues in the *Spiroplasma* sp. strain MQ1 DNA are methylated (1). Assuming that the enzyme responsible for this methylation is like many other prokaryotic DNA methylases, a *de novo* methylase, we undertook the cloning and expression in *E. coli* of the gene coding for this enzyme. We describe here the successful cloning and expression of the Spiroplasma gene in *E. coli* and its characterization. The structure of the Spiroplasma enzyme (M·SssI) is similar to that of all studied prokaryotic cytosine DNA methylases. Preliminary *in vitro* experiments indicate that the Spiroplasma methylase is a *de novo* methylase. This substrate specificity is characteristic of most prokaryotic DNA methylases. The enzyme (M·SssI) methylates completely and exclusively CpG sequences, a sequence specificity which is typical to all mammalian DNA methylases. It should be noted, however, that mammalian DNA methylases are primarily maintenance methylases methylating with higher efficiency hemimethylated DNA.

MATERIALS AND METHODS

Bacterial strains and plasmids

Spiroplasma sp. strain MQ1 was first isolated from a Monobia wasp and was provided by R.F. Whitcomb (U.S. Department of Agriculture, Beltsville, MD.) The *E.coli* strain ER 1451 (*mcr*A$^-$, *mcr* B$^-$, lacI^q) was obtained from New England Biolabs, Beverly, Mass. The plasmid pUC 18 was purchased from New England Biolabs. The pAL249 plasmid (derivative of pACYC 184) was kindly provided by A. Cohen (Department of Molecular Genetics, Hebrew University Medical School, Jerusalem, Israel). This plasmid was used to construct pSU77 by transferring the trp t176 gene, a strong tRNA opal suppressor under the control of *lac*UV5 promoter from pSWC101 to pAL 249(5). *E. coli* PR 101 was obtained by introducing a *rec*A mutation by P1 transduction of the strain ER 1451.

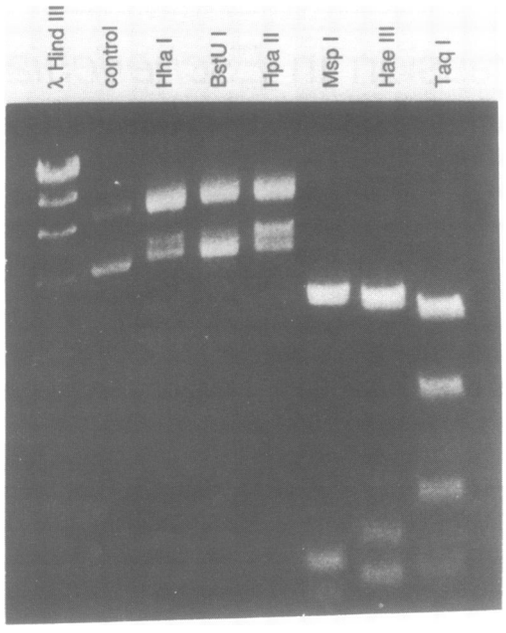

Figure 1. Restriction enzyme analysis of the positive clone pMT1. Plasmid DNA was prepared from a positive clone that was revealed in the preliminary screening. DNA samples were digested with the restriction enzymes as indicated in the figure using reaction conditions as recommended by the enzyme manufacturers. Undigested plasmid DNA was used as a control and HindIII fragments of λ phage DNA were used as size markers.

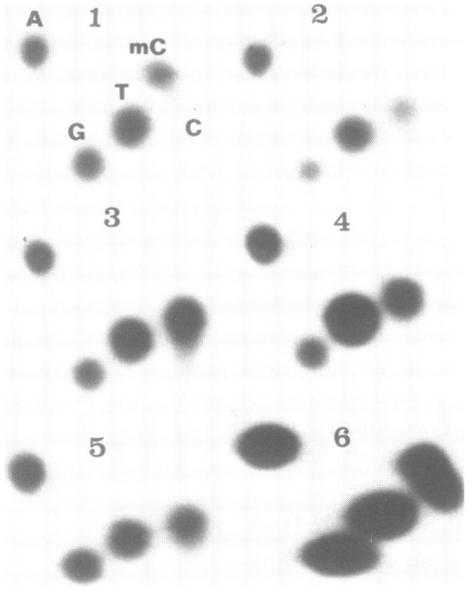

Figure 2. Nearest neighbor analysis of plasmid and chromosomal DNA from a positive clone. DNA was extracted from an overnight culture of PR 101 harboring pMT1/pSU77 after IPTG induction. The DNA was separated by CsCl ethidium-bromide isopycnic centrifugation,chromosomal DNA and plasmid DNA bands were isolated from the CsCl gradient and the ethidium-bromide removed. The two DNA samples were subjected to nearest neighbor analysis as described before (12). Plasmid DNA was nicked and end labelled, using 1-[α^{32}P]dGTP; 2-[α^{32}P]TTP; 3-[α^{32}P] dCTP; 4-[α^{32}P]dATP; E. coli chromosomal DNA was further purified from plasmid DNA on a 0.7% agarose gel, nicked and end labelled with [α^{32}P]dGTP: from 5-untransformed E. coli PR 101 cells and 6- the induced E. coli PR 101 culture harboring pMT1/pSU77. The labelled DNA was digested to 3′ nucleoside monophosphates that were separated by 2D TLC and autoradiographed as described (12). A, G, T, C, mC represent the 5′ nearest neighbors of the incorporated nucleotide.

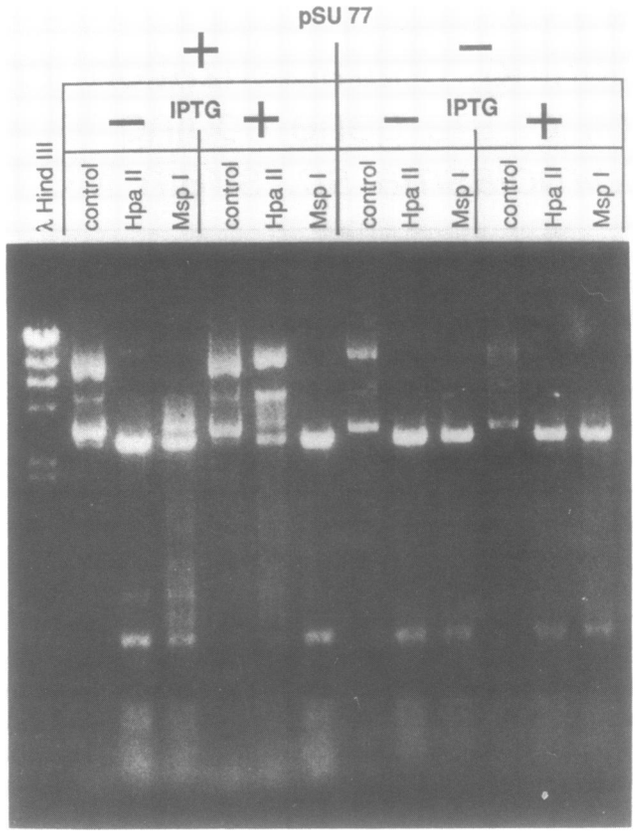

Figure 3. Requirement for the opal suppressor and IPTG induction. Minipreparations of plasmid DNA from cultures of *E. coli* PR 101 harboring pMT1 with and without the opal suppressor (pSU77) and grown in the presence or absence of 2.5 mM IPTG were digested with HpaII and MspI as recommended by the enzyme manufacturer. The digests were separated by agarose gel electrophoresis together with undigested plasmid DNA as controls and HindIII fragments of λ phage DNA as size markers.

Media

Spiroplasma sp. strain MQ1 was grown in Saglio medium (6) at 32°C. *E. coli* strains were kept on M9 plates supplemented with thiamine.

Transformations

All transformations were carried out according to Hanahan (7) without modification.

Construction of *Spiroplasma* sp. strain MQ1 genomic library

Spiroplasma sp. strain MQ1 DNA was partially digested by Sau 3A and 3−10 kb long fragments were pooled from a 10−40% sucrose gradient. The fragments were then ligated into the Bam H1 site of pUC18. The ligated DNA was used to transform *E. coli* PR 101 resulting in 5500 ampicillin resistant colonies. Restriction analysis of DNA minipreparations from individual colonies revealed that >95% of the colonies contain inserts of Spiroplasma DNA averaging in length about 5.5 kb.

Preparation of M·SssI methylase crude extract

A culture of *E. coli* PR101 harboring pMT1/pSU77 was induced with IPTG (2mM) at early log phase and grown to confluency. All steps of enzyme extraction were carried out at 0°−4°C. The cells were harvested and ground with twice their wet weight of alumina 305 (Sigma), and the paste was suspended in 2.5 volumes

of buffer A (10% glycerol [vol./vol.], 1mM dithiotheitol, 1mM EDTA, 40mM Tris-HCl pH8), this homogenous suspension was spun at 12,000×g for 10 min. Nucleic acids were removed by slow addition of 1/5 volume of 20% (wt./vol.) streptomycin sulfate (Nutritional Biochemical Co., Cleveland, Ohio) over a period of 30 min. The suspension was centrifuged for 10 minutes at 12,000×g, and 0.516 g ammonium sulfate (Merck) per ml supernatant (75% saturation) was added slowly over a period of 30 min. The ammonium sulphate precipitate was spun down, and resuspended in buffer A (10−20 μg protein/μl) and desalted on a Sephadex G-50 spin column before use. The crude enzyme (before desalting) has been kept at −70°C for over six months with no loss of activity.

DNA sequence analysis

Restriction fragments were subcloned into pBluescript KS(−) (Stratagene, Inc., San Diego, Cal.) Plasmids were purified by isopycnic centrifugation on CsCl gradients and used directly for double stranded sequencing by the dideoxy chain termination method with the modified T7 DNA polymerase, Sequenase (United States Biochemicals). The sequencing reactions were carried out according to the manufacturers protocol using [α^{35}S]dATP (Amersham, England) with primers purchased from Stratagene, Inc. Reactions were applied to 6% acrylamide denaturing gels and electrophoresis was performed at 2000V, 40mA for 3−4 hrs. The gels were fixed in 10% acetic acid, 12% methanol and dried. Autoradiography was carried out at room termperature for 16−72 hours using Agfa RP2 films.

RESULTS

The cloning strategy employed here was based on rendering the plasmid which carries the methylase gene, resistant to HhaI (GCGC) digestion by virtue of the methylation of CpGs in the plasmid DNA. This was expected to allow easy selection from a genomic library, of clones that express the methylase gene. This strategy, however, had to take into consideration several unpredictable situations. Firstly, it was possible that the E. coli transcription machinery would not recognize the Spiroplasma promoter. We therefore prepared the Spiroplasma. genomic library in pUC18 under the control of the *lac*Z promoter. Secondly, the methylation of CpGs in the plasmid or chromosomal DNA could render the DNA susceptible to restriction by the *E. coli mcr*A and *mcr*B methylcytosine restriction systems (8). For that reason, we used a *mcr*A$^-$, *mcr*B$^-$, *lac*I^q *E. coli* strain into which a *rec*A mutation had been introduced. This strain is designated here *E. coli* PR101 (see Methods). It was also possible that CpG methylation of the host DNA might interfere with the cell physiology. Therefore, we kept the *lac*Z promoter completely suppressed via *lac*I^q and induced the methylase by IPTG at the late logarithmic phase of culture growth. Thirdly, the universal stop codon UGA has been shown to code for tryptophan in *Mycoplasmas* (9, 10) and there are indications that the same situation exists in certain *Spiroplasma* species (11). To avoid premature termination of the biosynthesis of the methylase protein we constructed a compatible plasmid that carries the tRNA opal suppressor under the *lac*UV5 promoter, pSU77 (see Methods). Use of this construct enabled us to control the synthesis of the methylase by IPTG regardless of whether the transcription of the methylase gene is driven by its own promoter or the *lac*Z promoter, provided that the methylase gene contains at least one UGA triplet in its protein coding region.

Isolation of the MQ1 methylase gene

E. coli PR 101 cells harboring the opal suppressor pSU77 were transformed with the genomic library of *Spiroplasma* sp. strain MQ1 (constructed as described in the Methods). Ampicillin/chloramphenicol resistant colonies (about 20,000) were collected in LB medium. Aliquots of the cell suspension

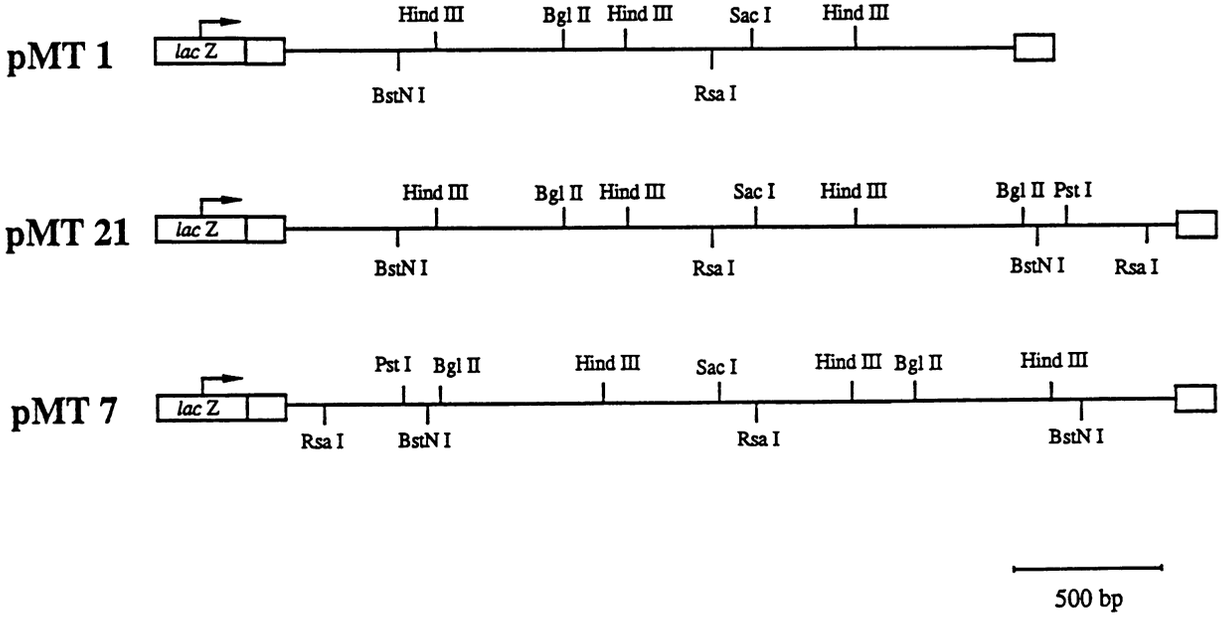

Figure 4. Partial restriction maps of the recombinant plasmids pMT1, pMT21 and pMT7. Restriction maps are presented of three clones which demonstrated methylase activity *in vivo* as described in figures 1 and 2. The Spiroplasma DNA fragments were inserted into the BamH1 site of pUC18. The direction of the *lac*Z promoter in relation to the insert orientation is designated by the arrow.

were used to grow cultures to 10^8 cells/ml. IPTG was added to a concentration of 2.5 mM and the culture grown to confluency, plasmid DNA was prepared, and 10 μg was digested to completion with 25U HhaI. The digested DNA was used for a second transformation of PR 101/pSU77. Sixty-six amp^R/cam^R colonies were obtained, while a parallel transformation with undigested DNA yielded 10^7 transformants/μg DNA.

Individual cultures of each of the surviving colonies were induced by IPTG and minipreparations of plasmid DNA were used to analyze CpG methylation by HpaII/MspI digestions. DNAs of approximately 60% (44 out of 66) of the clones were resistant to digestion by HpaII (CCGG) and by other CpG methyl sensitive enzymes (HhaI, Bst UI [CGCG]) while these DNAs were readily digested by enzymes that cut at CpG containing sites regardless of their methylation state. (MspI [CCGG], Taq I [TCGA]). As an example of this analysis, the restriction digests of one clone are presented in Fig. 1. Plasmids from positive clones, as judged by the restriction enzyme analysis, were designated pMT.

Plasmid and chromosomal DNA from several positive clones were subjected to further analysis of their CpG methylation state using the nearest neighbor method (12). This analysis provides both a quantitative estimate of the extent of methylation at CpG sequences and a convenient tool for examination of sequence specificity of the methylase. The results of such an analysis presented in Fig. 2, clearly demonstrate that the cloned enzyme methylates exclusively and completely CpG sequences in the plasmid DNA. A 5-methylcytosine (mC) spot appears only when DNA is labelled with [α^{32}P] dGTP (Panel 1). No such spot is detected when DNA is labelled with either dTTP (Panel 2), dCTP (Panel 3), or dATP (Panel 4). Interestingly, chromosomal DNA in the same cells was only 50–60% methylated (Panel 6). Panel 5 shows chromosomal DNA of non transformed cells labelled with dGTP.

Characterization of the methylase gene

To examine the requirement for opal suppression and the effect of IPTG, PR 101 cells harboring pMT1 with and without pSU77 (the opal suppressor) were grown to 10^8 cells/ml. Each culture was divided and IPTG was added (to 2.5 mM) to one portion. DNA minipreparations of all four combinations were subjected to HpaII/MspI analysis. The results reveal a requirement for both IPTG induction and the presence of the opal suppressor to express the methylase (Fig. 3). The requirement for IPTG raises the question of whether the methylase gene is induced directly by IPTG or indirectly via induction of the opal suppressor. In an attempt to answer this question, restriction maps of three representative positive clones were constructed. This analysis

A

```
1     GATCTTGATAATTCAACTTTATTTGGATGTTGTGAAATGTAGATAAATCTTACATTTTTT
                    ↓                          M  S  K  V  E  N     6
61    ATTGTAAAATCTTATTAGATTGTTTTTTTAGAGGAGCTAAAAATGAGCAAAGTAGAAAAT
      K  T  K  K  L  R  V  F  E  A  F  A  G  I  G  A  Q  R  K  A    26
121   AAAACAAAAAAACTTAGAGTATTTGAAGCTTTTGCTGGAATTGGTGCTCAAAGAAAAGCC
      L  E  K  V  R  K  D  E  Y  E  I  V  G  L  A  E  W  Y  V  P    46
181   TTGGAGAAAGTCAGAAAAGATGAATATGAAATAGTAGGGCTTGCTGAATGATATGTTCCT
      A  I  V  M  Y  Q  A  I  H  N  N  F  H  T  K  L  E  Y  K  S    66
241   GCAATTGTTATGTATCAAGCTATACACAACAATTTTCATACAAAGTTGGAGTATAAATCA
      V  S  R  E  E  M  I  D  Y  L  E  N  K  T  L  S  W  N  S  K    86
301   GTTTCTAGAGAAGAAATGATTGACTATTTGGAAAATAAAACACTATCTTGAAACTCAAAA
      N  P  V  S  N  G  Y  W  K  R  K  K  D  D  E  L  K  I  I  Y   106
361   AATCCAGTATCTAATGGTTATTGGAAGAGAAAAAAAGATGATGAACTTAAAATTATATAT
      N  A  I  K  L  S  E  K  E  G  N  I  F  D  I  R  D  L  Y  K   126
421   AATGCAATTAAGTTATCTGAAAAAGAGGGTAATATTTTTGATATTAGAGACCTTTACAAA
      R  T  L  K  N  I  D  L  L  T  Y  S  F  P  C  Q  D  L  S  Q   146
481   AGAACTTTGAAAAATATAGATTTATTAACATATTCATTTCCTTGTCAAGACTTATCTCAA
      Q  G  I  Q  K  G  M  K  R  G  S  G  T  R  S  G  L  L  W  E   166
541   CAGGGTATTCAAAAGGGTATGAAAAGAGGTTCTGGTACTAGATCAGGTCTCTTATGAGAA
      I  E  R  A  L  D  S  T  E  K  N  D  L  P  K  Y  L  L  M  E   186
601   ATTGAAAGAGCTTTGGATTCAACTGAAAAAAATGACTTACCAAAATACTTGTTAATGGAA
      N  V  G  A  L  L  H  K  K  N  E  E  E  L  N  Q  W  K  Q  K   206
661   AATGTAGGAGCTCTTCTTCACAAGAAGAATGAAGAAGAACTAAATCAATGAAAGCAAAAA
      L  E  S  L  G  Y  Q  N  S  I  E  V  L  N  A  A  D  F  G  S   226
721   TTAGAAAGTCTTGGCTATCAAAACTCAATTGAAGTTTTGAATGCCGCTGACTTCGGTTCC
      S  Q  A  R  R  R  V  F  M  I  S  T  L  N  E  F  V  E  L  P   246
781   TCACAAGCAAGAAGAAGAGTTTTTATGATATCTACTTTAAATGAATTTGTTGAACTACCA
      K  G  D  K  K  P  K  S  I  K  K  V  L  N  K  I  V  S  E  K   266
841   AAGGGAGATAAAAAACCTAAAAGTATCAAAAAAGTTTTAAATAAAATAGTTTCTGAAAAA
      D  I  L  N  N  L  L  K  Y  N  L  T  E  F  K  K  T  K  S  N   286
901   GATATTTTAAATAATTTATTGAAATATAATTTAACTGAATTTAAAAAAACAAAATCAAAT
      I  N  K  A  S  L  I  G  Y  S  K  F  N  S  E  G  Y  V  Y  D   306
961   ATAAATAAAGCTTCACTGATTGGTTACAGTAAATTTAATTCAGAAGGTTATGTTTATGAT
      P  E  F  T  G  P  T  L  T  A  S  G  A  N  S  R  I  K  I  K   326
1021  CCTGAATTTACAGGACCAACCTTAACTGCAAGCGGTGCAAATTCAAGAATAAAAATCAAA
      D  G  S  N  I  R  K  M  N  S  D  E  T  F  L  Y  I  G  F  D   346
1081  GATGGATCTAATATTAGAAAAATGAATTCAGACGAAACTTTCTTATATATTGGGTTTGAT
      S  Q  D  G  K  R  V  N  E  I  E  F  L  T  E  N  Q  K  I  F   366
1141  TCACAAGATGGAAAAAGAGTAAATGAAATTGAATTTTTAACTGAAAATCAAAAAATATTT
      V  C  G  N  S  I  S  V  E  V  L  E  A  I  I  D  K  I  G  G   386
1201  GTTTGTGGAAATTCAATATCAGTAGAAGTTTTGGAAGCGATTATAGATAAAATTGGAGGT
      *
1261  TAATACATGTCAATAAAAGATAAAACATATTTAGTAAATGTTCCTATAATAAGAAAAAAT
1321  TTTTTTTCTACTAATTCAGCTAAGTTTAAGGGAAATTTGCAAATTATTTCAAATGATTTT
1381  GGAATAATAAAAGATTGAAAAATTGAATTCAAAGAAGTTTGAATAGATATATTTACAACC
1441  CCAAGGTGATTTAGTGATAAGCCTAGCTTATTTAAATATAGGGAAAGGCTTTTTGAAGAA
1501  TTAAATTTAATTCAGGAGCACAGACCTTCAAAAAGATC 1538
```

B

```
M·HhaI      1  MIEIKDKQLTGLRFIDLFAGLGGFRLALESCGA...ECVYSNEWDKYAQEVYEMNFGEKPE.....................................
M·SssI      1   MSKVENKTKKLRVFEAFAGIGAQRKALEKVRKDEYEIVGLAEWYVPAIVMYQAIHNNFHTKLEYKSVSREEMIDYLENKTLSWNSKNPVSNGYWKRKKDDELKI
SPR         1          MGKLRVMSLFSGIGAFEAALRNIGV.GYELVGFSEIDKYAVKSFCAIHNVDEQLNF.....................................

Consensus       ---------------SLFSG-GG---G-E--G-   ------E-DK-A--TY--N--------

                                Region I                Region II


M·HhaI     59  .............GDITQVNEKTIPDHDILCAGFPCQAFSISGKQKGFE...DSRGTLFFDI....ARIVREKKPKVVFMENVKNFASHDNGNTLEVVKNTMNE
M·SssI    105  IYNAIKLSEKEGNIFDIRDLYKRTLKNIDLLTYSFPCQDLSQQGIQKGMKRGSGTRSGLLWEIERALDSTEKNDLPKYLLMENVGALLHKKNEEELNQWKQKLES
SPR        56  .............GDVSKIDKKKLPEFDLLVGGSPCQSFSVAGHRKGFE...DTRGTLFFQY....VETLKEKQPKFFVFENVKGLINHDKGNTLNVMAEAFSE

Consensus                    GDI--------P--D-L--GFPCQ-FS-AG-R-G--   D-RG-LFF--    -R------PK-F--ENVKGL-----G--L--I------

                         Region III             Region IV          Region V              Region VI


M·HhaI     143  LDYSFHAKVLNALDYGIPQKRERIYMICFRNDLNIQ ~ TRKLHPRECARVMGYPDSYK......VHPSTSQAYKQFGNSVVINVLQYIAYNIGSSLNFKPY*   307
M·SssI     210  LGYQNSIEVLNAADFGSSQARRRVFMISTLNEFVEL ~ IRKMNSDETFLYIGFDSQDGKRVNEIEFLTENQKIFVCGNSISVEVLEAIIDKIGG*          386
SPR        140  VGYRIDLELLNSKFFNVPQNRERLYIIGIREDLIKN ~ IRRLTPLECFRLQAFDDEDFEKA.FAAGISNSQLYKQAGNSITVTVLESIFKELIHTYVNKESE*  439

Consensus       -GY-V----LNA---GVPQ-RER--IIG-R-D---N ~ -R-LT--E-AR-QGFPD------ -----S--Q-YKQ-GNSV-V----AIA--I----------

           Region VII     Region VIII                    Region IX                 Region X
```

C

Methylase	Recognition sequence	Position	Amino acid sequence
Mouse	CG	1376-1418	TLGTGTTTGLASTGVWSGMASSAP**LSPTLS**PWQAGSGAPPGAA
M·*Sin*I	GG(A/T)CC	336-378	EALGKSFFLGGGKTGFLRRV**A**WDRP**SPTLVT**HPAMPATDLAHP
ρ11s (TRD$_2$)	?	309-349	DPIMIGHIDLKGHDAIKRV**Y**SPD**GVSPTLTT**MGEGHREPKI..
ϕ3T (TRD$_2$)	GC(N)GC	288-330	EPLMVGHVDLKGHDAIKRV**Y**SPE**GLSPTLTT**MGGGHREPKIAE
SPR (TRD$_3$)	CC(A/T)GG	253-295	TVEQVGNINPSGNGMNGN**VYN**SS**GLSPTITT**NKGEGLKIAVEY
ρ11s (TRD$_3$)	?	350-392	AVEYVGNINPSGKGMNDQG**Y**NSN**GLSPTLTT**NKGEGVKISVPN
M·*Sss*I	CG	286-328	NINKASLIGYSKFNSEGY**VY**DPEFTG**PTLT**ASGANSRIKIKDG
ρ11s (TRD$_1$)	G(N)GC(N)C	262-304	RWLNNGKTYSRNFKQGNR**VY**DSNGIA**TTLTS**QSVGGLGGQTSL
M·*Hha*I	GCGC	223-265	TPKTVRLGIVGKGGQGER**IY**STRG**IAITLS**AYGGGIFAKTGGY

Figure 5. Nucleotide and animo acid sequence analysis of M·SssI. A. Nucleotide sequence of the 1538-bp BglII−PstI fragment (PstI site from the polylinker of pUC18, see Fig. 4) containing the M·SssI gene. The gene is presumed to start at the ATG codon at position 103 and to stop at the TAA codon at position 1261. An arrow marks the transcription start point. The putative ribosome binding site is shown in bold-face type, and the putative −35 and −10 regions are underlined. Nucleotide numbering is given at the left of each row. The derived amino acid sequence for the M·SssI methylase is shown using the single-letter amino acid code. Amino acid numbering is given at the right of each row. The four tryptophan residues encoded by TGA codons are shown italicised (*W*) . The amino acid sequence was derived using the University of Wisconsin Genetics Computer Group (UWGCG) program 'map'. B. Amino acid sequence alignments between M·SssI and its two closest relatives, M·HhaI and SPR. Gaps introduced into the sequences of M·HhaI and SPR to maximize alignment are indicated by dots (·). Amino acid numbering is shown at the beginning of each line. For simplicity, the 'variable' region of each protein, between conserved regions VIII and IX, has been omitted from the figure. The ommissions are signified in the figure by a curly dash (~). The consensus sequence below the alignments is based upon a compilation of nineteen sequenced cytosine DNA methylases: the thirteen sequences of Posfai et al. (15), and several recent additions (16, 27); and M·*Sss*I (this paper). The sequences were aligned by eye using UWGCG 'Lineup', and the consensus was calculated using UWGCG 'Pretty'. Inclusion of an amino acid in the consensus sequence indicates that the amino acid occurs at that position in more than 50% of cytosine DNA methylases. Symbols printed in **bold** type occur at that position in over 90% of cytosine DNA methylases. (Bold type has also been used to highlight the corresponding amino acids in the M·HhaI, M·SssI and SPR sequences). A dash (−) in the consensus indicates that an amino acid occurs at that position in most sequences, but that no single amino acid makes up a majority. A blank in the consensus indicates that only a minority of sequences include an amino acid at that position. The ten regions of sequence similarity defined by Posfai et al. (15) are shown below the consensus. C. Partially conserved sequence motif within the TRD's of the *Bacillus* phage cytosine DNA methylases and the variable regions of M·SssI, M·SinI, M·HhaI and the mouse MTase. Conserved amino acids are shown in **bold** type. Related amino acids were grouped as described before (14): (E, D,Q, N), (L, I, V,F, M, C, Y, A, W), (H, R, K), (S, T), (G, P).

revealed two different clones with identical (3kb) inserts in opposite orientations (Fig. 4). This observation clearly indicates that in at least one of these clones transcription of the methylase gene is driven by an endogenous promoter. This conclusion was confirmed by primer extension experiments (data not shown). In light of this fact, along with the observed requirement for IPTG induction and opal suppression, we have concluded that the effect of IPTG on the production of the methylase is brought about through induction of the opal suppressor gene.

Sequencing of the M·SssI gene

A 1538-bp BglII−PstI fragment from pMT1 (see fig. 4) was subcloned and shown to specify the DNA methylase. The fragment was sequenced on both strands by the dideoxy chain-termination method, and the amino acid sequences of the predicted translational products from each reading frame were examined. The ATG codon at position 103 was tentatively identified as the start of the methylase gene. An open reading

frame (ORF) beginning at position 103 is 1158 bp in length, and includes four TGA codons at positions 229, 349, 595 and 709, which we assume code for tryptophan (Fig. 5A). This ORF encodes a 386 amino acid protein that contains all of the conserved sequence motifs characteristic of cytosine DNA methylases, including the invariant Pro-Cys dipeptide that is thought to constitute the active site for transmethylation (Fig. 5B) (13, 14, 15). Primer extension analysis has identified position 77 on the DNA sequence to be the start point of transcription (data not shown). A putative ribosome binding site sequence AGGAG, preceeds the start codon by 6 nucleotide residues. Another ORF starts at position 1267 of this sequence. Further investigation is required to clarify if this ORF has a biological function.

Amino acid sequence comparisons

The sequences of eighteen bacterial cytosine DNA methylases that are presently known have been compared and found to closely resemble one another; in general, they span 300 to 400 amino acid residues in length and contain ten, or so, regions of amino acid homology (14, 15). The deduced amino acid sequence for M·SssI is similar to those of the other cytosine DNA methylases; it displays greatest similarity to M·HhaI and SPR (Fig. 5b), and least similarity to M·BspRI. The first conserved region in M·SssI lies close to the N-terminus, extending from positions13 to 32. The remaining conserved regions comprise amino acid residues 37−55 (region II); 116−120 (region III); 132−155 (IV); 159−166 (V); 179−198 (VI); 210−215 (VII); 219−237 (VIII); 331−347 (IX); and 359−382 (X). The most striking difference between M·SssI and other cytosine DNA methylases occurs between conserved regions II and III. In most methylases this interval is approximately 10 amino acids in length, but in M·SssI it is 60 residues in length. Only one other methylase, M·BepI (16), contains a large insert at this position; the insert in M·BepI is approximately 40 amino acids long, and bears no similarity to that of M·SssI.

The interval between conserved regions VIII and IX in cytosine DNA methylases forms the so-called 'variable region'. The length and sequence of this region varies considerably from enzyme to enzyme. In M·SssI, the region is 85 residues in length and it extends from position 246 to 330. The corresponding variable regions in the *Bacillus* phage methylases, ϕ3T, r11s and SPR (17−20), have been shown to be responsible for DNA sequence-specificity (21−23). Since all cytosine DNA methylases appear to have the same basic organization, it is assumed that sequence-specificity in all of these enzymes is determined by the variable region. In support of this idea is the finding that enzymes that perform identical reactions often possess closely similar variable regions, whereas enzymes that carry out different reactions possess different variable regions. This holds true for M·BspRI, M·BsuRI, M·HaeIII and M·NgoPII, which produce GG^{m}CC (24−27). Besides M·SssI, only one other cytosine DNA−methylase that recognizes CpG, the mouse DNA methylase has been cloned and characterized (28). Comparison between the variable region of the mouse methylase and that of M·SssI (both recognize and methylate exclusively CpG sequences) does not reveal the extensive similarities that are seen among the isomeric methylases mentioned above.

The variable regions in the *Bacillus* phage methylases that methylate various DNA sequences, are the best understood. They consist of an initial section of approximately 75 amino acids followed by several 'target recognition domains' (TRDs), each consisting of approximately 50 amino acids in length (22). Lauster et al. (14) have noted that the TRDs of the *Bacillus* phage methylase bear a partial resemblance to one another. The sequence motif that is partially conserved in the TRDs also occurs in the variable regions of several other cytosine DNA methylases (14), including the mouse DNA methylase, M·HhaI, M·SinI (28−30) and the enzyme described here, M·SssI. An alignment of this motif in these regions is shown in Fig. 5c. The presence of the motif does not correlate in any simple way with the recognition of a common DNA sequence. Whereas the mouse methylase, M·SssI and M·HhaI produce ^{m}CG in DNA, the reactions performed by the *Bacillus* phage methylases do not create this product. Furthermore, the motif is not observed in the variable regions of other methylases that do create ^{m}CG, such as M·BepI (product: ^{m}CGCG) (16).

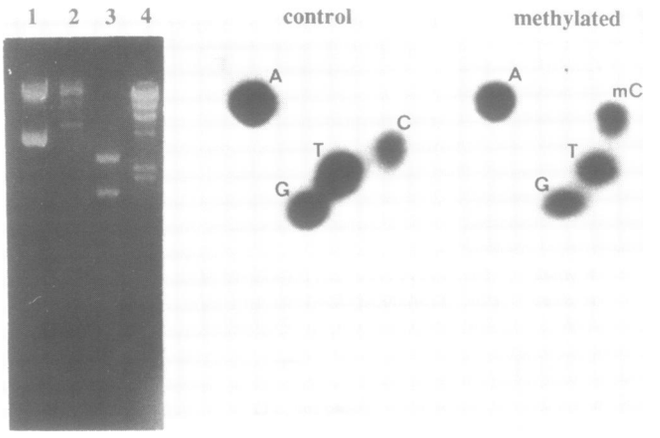

Figure 6. Substrate and Sequence specificity of M·SssI *in vitro*. Crude extracts of PR 101 cells harboring pMT100 /pSU77 induced with IPTG (see Materials and Methods) were used to methylate ϕX174 RF DNA in vitro. 20 μg protein was incubated in a standard methylation reaction mixture containing 50 mM Tris-HCl pH 8, 10 mM EDTA, 5 mM DTT, 160μM SAM, and 0.1 mg/ml bovine serum albumin, at 28°C for 18 hrs. At the end of incubation plasmid DNA was purified by one phenol/ chloroform extraction and subjected to digestion by HpaII. Undigested DNA Control (lane 1), methylated DNA digested by HpaII (lane 2), unmethylated DNA digested by HpaII (lane 3), Hind III fragments of λ phage DNA (lane 4). Nearest neighbor analysis was carried out on unmethylated and *in vitro* M·SssI methylated ϕx DNA. Procedure and spots on the autoradiographs are as described in the legend to Fig. 2.

Characterization of the methylase

A crude extract of the cells harboring pMT1 and grown in the presence of IPTG was treated with streptomycin sulfate to remove nucleic acids and precipitated with ammonium sulfate(75% saturation) at 0°C (Materials and Methods),(31). This enzyme preparation has been used for *in vitro* methylation of ϕX174 DNA. The enzyme methylated completely and exclusively CpG sequences in ϕX DNA as judged by restriction enzyme analysis and by the nearest neighbor method (Fig. 6). In an experiment in which substrate specificity was examined, unmethylated DNA was compared with hemimethylated DNA as substrates. Both substrates were methylated with the same efficiency (data not shown), clearly indicating that the methylase is a *de novo* methylating enzyme. A more detailed study concerning the characterization of the methylase is underway.

DISCUSSION

The DNA methylase of *Spiroplasma* sp. strain MQ1 displays the same sequence specificity as that of the known mammalian DNA methylases (1). The sequence specificity of this methylase is

characterized by methylating exclusively CpG sequences, making it a convenient tool for studying DNA methylation in eukaryotes. *Spiroplasma* sp. strain MQ1 cells grow 10 fold slower than *E. coli* and reach confluency at a very low cell density (10^8 cells/ml) in a complex and expensive medium. These facts and our interest in eukaryotic DNA methylation prompted us to undertake the cloning of the MQ1 methylase gene and express it in *E. coli*. We have described here the successful cloning and expression of the gene coding for the Spiroplasma DNA methylase in *E. coli*.

While cloning the Spiroplasma methylase gene several interesting observations have been made. Firstly, although the gene was isolated from a genomic library in pUC18 with the intention that it would be controlled by the *lac*Z promoter, two different clones turned out to have the Spiroplasma DNA insert in opposite orientations. This implies that in at least one of these clones the Spiroplasma promoter is recognized by the *E. coli* transcriptional apparatus and drives the transcription of the gene. This was verified by primer extension experiments which demonstrated that the start point of transcription was in the spiroplasma insert. Secondly, it has been suggested that UGA may be read in *Mollicutes* as a tryptophan codon (9,11). The observed absolute requirement for an opal suppressor to express the methylase gene implied that in fact, UGA does code for tryptophan in *Spiroplasma* sp., and that the methylase gene contains at least one TGA codon. Indeed, the sequence analysis presented here revealed four TGA codons presumably read as tryptophan. This fact enabled us to control the potentially lethal methylase activity within the cell by placing the TrpT176 opal suppressor under the control of the *lac*UV5 promoter and inducing synthesis of the suppressor with IPTG. However, induction of the methylase by IPTG at the time of innoculation had only a limited effect on the growth of the cells. This suggests that methylation at CpGs of the chromosomal DNA does not interfere with the cell physiology. In this regard it should be mentioned that a full length cDNA of the mouse DNA methylase that has recently been cloned (28) has not yet been reported to express in *E. coli*. It should be noted, however, that the chromosomal DNA was methylated only to the extent of about 50%. It is, therefore, possible that the system selects against cells that are methylated at all CpGs in the chromosomal DNA.

Although recombinant DNA technology has made it possible to clone and express Mollicute DNA fragments in *E. coli*, in most cases the Mollicute DNA was not fully expressed. This is explained by the observation that the universal stop codon UGA is read in the *Mollicutes* as tryptophan (9,11). Only Spiralin, a membrane protein of *Spiroplasma citri* was fully expressed in *E. coli* and shown subsequently to have no tryptophan residues (11).The methylase gene described here, is therefore, the first Spiroplasma gene (containing a TGA codon) to be cloned and successfully expressed in *E. coli*. The PR 101 /pSU77 system has proved suitable for future cloning and expression of Spiroplasma and mitochondrial genes in *E. coli*.

M·SssI, a CpG methylase is a unique prokaryotic DNA methylase, firstly, in that it has a dinucleotide recognition sequence, and secondly, in that it shares this recognition sequence with the eukaryotic DNA methylase. *Spiroplasma* is a plant parasite and it is possible that this bacteria inherited the gene from it's eukaryotic host. It was, therefore, interesting to examine whether M·SssI and the mouse DNA methylase share structural features in addition to their substrate sequence specificity. Comparison of the M·SssI deduced amino acid sequence and that of the mammalian enzyme revealed that the two enzymes are only distantly related. As a whole, M·SssI shows strong homology to all of the conserved regions of the prokaryotic cytosine DNA methylases, whereas homology to it's isoschizomer the mouse maintenance methylase is limited to overall similarities seen between the prokaryotic and eukaryotic methylases in general. Additionally, preliminary results have demonstrated that M·SssI, unlike it's eukaryotic counterpart, is a *de novo* methylase, being equally active on unmethylated duplex DNA as it is on hemimethylated duplex DNA, a typically prokaryotic trait.

ACKNOWLEDGMENTS

We are grateful to H. Cedar for careful reading of the manuscript and useful suggestions. This study was partially supported by NIH grant GM20483 to A. R.

REFERENCES

1. Nur, I., Szyf, M., Razin, A., Glaser, G. , Rottem, S. and Razin, S. (1985) J. Bacteriol **164,** 19–24.
2. Razin, A. and Riggs, A. D. (1980) Science **210**, 604–610.
3. Razin, A. and Cedar, H. (1984) Int. Rev. Cytobiol. **92,** 159–185.
4. Razin, A. (1989) In Setlow, J.K. (ed.), Genetic Engineering, Plenum Press, N.Y. Vol. 11, pp. 1–12.
5. Raftery, L.A., Egan, J.B., Cline, S.W. and Yarus, M. (1984) J. Bacteiol **158,** 849–859.
6. Saglio, P., Lafleche, D., Bonissol, C. and Bove, J.M. (1971) Physio. Veg. **9,** 569–582.
7. Hanahan, D. (1985) In Glover, D.M. (ed.), DNA Cloning-A Practical Approach, IRL Press, Oxford, Vol.I, pp 109–135.
8. Raleigh, E.A. and Wilson, G. (1986) Proc. Natl. Acad. Sci. U.S.A. **83,** 9070–9074.
9. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. and Osawa, S. (1985) Proc. Nat. Adad. Sci. (U.S.A.) **82,** 2306–2309.
10. Inamine, J.M., Denny, T.P., Loechel, S., Shaper, U., Huang, C., Bott, K.F. and Hu, P. (1988) Gene **64,** 217–229.
11. Bove, J.M., Carle, P., Garnier, M., Laigret , F., Renandim, J. and Saillard, C., (1988) In Whitcomb, R.F. and Tully, J.G. (eds.), The Mycoplasmas, Academic Press, New York, Vol. 5, pp 244–364.
12. Gruenbaum, Y., Stein, R., Cedar, H. and Razin, A. (1981) FEBS letters **142,** 67–71.
13. Wu, J.C. and Santi, D. V. (1987) J. Biol. Chem. **262,** 4778–4786.
14. Lauster, R., Trautner, T.A. & Noyer-Weidner, M. (1989) J. Mol. Biol. **206,** 305–312.
15. Posfai, J., Bhagwat, A. S., Posfai, G., and Roberts, R. J. (1989) Nucleic Acids Res. **17,** 2421–2435.
16. Kupper, D., Zhou, J-G., Venetianer, P. and Kiss, A. (1989) Nucleic Acids Res. **17,** 1077–1088.
17. Buhk, H.J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Gunthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) Gene **29,** 51–61.
18. Posfai, G., Baldauf, F., Erdei, S., Posfai, J., Venetianer, P. and Kiss, A. (1984) Nucleic Acids Res. **12,** 9039–9049.
19. Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. and Trautner, T.A. (1986) Gene **42,** 89–96.
20. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesh, T.S. and Trautner, T.A. (1987) EMBO J. **6,** 1137–1142.
21. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. & Trautner, T.A. (1988) EMBO J., **7,** 2601–2609.
22. Trautner, T.A., Balganesh, T.S. & Pawlek, B. (1988) Nucleic Acids Res. **16,** 6649–6658.
23. Balganesh, T.S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K. and Trautner, T.A. (1987) EMBO J. **6,** 3543–3549.
24. Posfai, G., Kiss, A., Erdei, S., Posfai, J. and Venetianer, P. (1983) J. Mol. Biol. **170,** 597–610.33.
25. Kiss, A., Posfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) Nucleic Acids Res. **13,** 6403–6421.
26. Slatko, B.E., Croft, R., Moran, L. and Wilson, G.G. (1988) Gene 74, 45–50.
27. Sullivan, K.M. and Saunders, J.R. (1988) Nucleic Acids Res. **16,** 4369–4387.

28. Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988) J. Mol. Biol. **203,** 971–983.
29. Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G.G. and Wells, R.D. (1987) J. Biol. Chem. **262,** 4770–4777.
30. Karreman, C. and de Waard, A. (1988) J. Bacteriol **170,** 2527–2532.
31. Urieli-Shuval, S., Gruenbaum, Y. and Razin, A. (1983) J. Bacteriol. **153,** 274–280.